PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-202790
(43)Date of publication of application : 19.07.2002

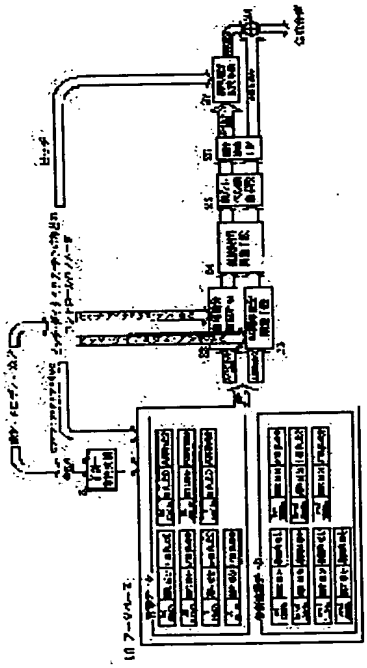
(51)Int.CI. G10L 13/06
G10L 13/08
G10L 13/00

(21)Application number : 2000-401041
(22)Date of filing : 28.12.2000

(71)Applicant : YAMAHA CORP
(72)Inventor : KENMOCHI HIDENORI
XAVIER SERA
JORDI BONADA

(54) SINGING SYNTHESIZER

(57)Abstract:
PROBLEM TO BE SOLVED: To synthesize singing voices of high quality.
SOLUTION: A spectrum model synthesis(SMS), which is an analytical and synthetic process, is conducted about the phoneme or two or more phoneme chains, a database 10 is prepared, and the SMS data of the phoneme or phoneme chains required for synthesis are concatenated and synthesized, to obtain singing voices. Into the database 10, separate segment data are stored by each different pitch, dynamics, and tempo concerning the same phoneme or phoneme chain. A harmonic component adjustment means 22 and a non-harmonic component adjustment means 23 adjust the harmonic components and non-harmonic components of read segment data so as to match them to a target pitch. A duration adjustment means 24 adjusts the length of the phonemes or the phoneme chains with the length matching the target tempo. A segment level adjustment means 25 carries out level adjustment, and then connects individual segments, generates harmonic components corresponding to the desired pitch, and synthesizes high quality singing voices with a non-harmonic components and the harmonic component.



LEGAL STATUS
[Date of request for examination] 16.10.2001

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2002－202790
（Ｐ２００２－２０２７９０Ａ）
(43)公開日　平成14年７月19日(2002.7.19)

(51)Int.Cl.[7]	識別記号	ＦＩ		テーマコード（参考）
G10L 13/06		G10L 5/04	F	5D045
13/08		3/00	H	
13/00		5/04	D	
		7/00	C	

審査請求　有　請求項の数13　ＯＬ　（全17頁）

(21)出願番号　特願2000－401041(Ｐ2000－401041)

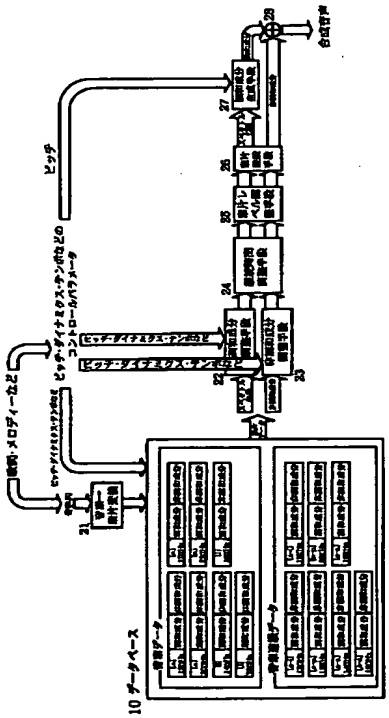
(22)出願日　平成12年12月28日(2000.12.28)

(71)出願人　000004075
ヤマハ株式会社
静岡県浜松市中沢町10番１号
(72)発明者　劔持　秀紀
静岡県浜松市中沢町10番１号　ヤマハ株式会社内
(72)発明者　ザビエル　セラ
スペイン　バルセロナ　08002　メルセ　１２
(74)代理人　100102635
弁理士　浅見　保男　（外２名）

最終頁に続く

(54)【発明の名称】歌唱合成装置

(57)【要約】
【課題】　高品質な歌声を合成する。
【解決手段】　スペクトルモデル合成（ＳＭＳ）分析合成法において、音素または２つ以上の音素連鎖についてＳＭＳ分析を行いデータベース１０を作成し、合成時に必要な音素または音素連鎖のＳＭＳデータを接続し合成することで歌声を得る。前記データベース１０には、同じ音素あるいは音素連鎖につき、異なるピッチ、ダイナミクス、テンポごとに別個の素片データを記憶する。調和成分調整手段２２、非調和成分調整手段２３で、読み出した素片データの調和成分および非調和成分を目的のピッチに合うように調整し、継続時間調整手段２４で目的のテンポに合うように音素または音素連鎖の長さを調整し、素片レベル調整手段２５でレベル調整した後、各素片を接続し、所望のピッチに対応した調和成分を生成して、非調和成分と合成する。



1

【特許請求の範囲】

【請求項１】　音素あるいは２つ以上の音素のつながりである音素連鎖である音声素片について調和成分のデータと非調和成分のデータを記憶した音韻データベースを有し、歌詞に対応した音声素片データを前記音韻データベースから読み出して接続することにより、歌唱音を合成する歌唱合成装置であって、

目的のテンポや歌い方に合うように前記音韻データベースから読み出した音声素片データの時間長を調整する継続時間調整手段と、

目的のピッチに合うように前記音韻データベースから読み出した音声素片データの前記調和成分および前記非調和成分を調整する調整手段とを有することを特徴とする歌唱合成装置。

【請求項２】　前記音声素片データを接続するときに、調和成分、非調和成分それぞれについてスムージング処理あるいはレベル調整処理を行なう素片レベル調整手段を有することを特徴とする請求項１記載の歌唱合成装置。

【請求項３】　前記音韻データベース中には、同一の音素または音素連鎖について、ピッチ、ダイナミクス、テンポの異なる複数の音声素片データが記憶されていることを特徴とする請求項１あるいは２記載の歌唱合成装置。

【請求項４】　前記音韻データベース中には、母音などの伸ばし音からなる音声素片データ、子音から母音あるいは母音から子音への音素連鎖からなる音声素片データ、子音から子音への音素連鎖からなる音声素片データおよび母音から母音への音素連鎖からなる音声素片データが記憶されていることを特徴とする請求項１～３のいずれかに記載の歌唱合成装置。

【請求項５】　前記調和成分のデータと前記非調和成分のデータは、その素片の区間に含まれるフレーム列の各フレームに対応する周波数領域のデータ列として記憶されていることを特徴とする請求項１～４のいずれかに記載の歌唱合成装置。

【請求項６】　前記継続時間調整手段は、音声素片に含まれるフレーム列中の１または複数のフレームを繰り返すこと、あるいは、フレームを間引くことにより所望の時間長のフレーム列を生成するものであることを特徴とする請求項５記載の歌唱合成装置。

【請求項７】　前記継続時間調整手段は、非調和成分のフレームを繰り返すときに、合成時に時間的に逆行する場合には、その非調和成分の位相スペクトルの位相を反転させることを特徴とする請求項６記載の歌唱合成装置。

【請求項８】　歌唱音合成時に、調和成分について、音声素片データに含まれている調和成分のスペクトル包絡の概形を保ったままピッチだけを所望のピッチに変換する調和成分生成手段を有することを特徴とする請求項５記載の歌唱合成装置。

2

【請求項９】　前記音韻データベース中に記憶される音声素片データのうち伸ばし音に対応する音声素片については、非調和成分の振幅スペクトルとして、その非調和成分の振幅スペクトルにその伸ばし音の区間を代表するスペクトルの逆数を乗算することにより得られた平坦なスペクトルを記憶していることを特徴とする請求項５記載の歌唱合成装置。

【請求項１０】　歌唱音合成時に、伸ばし音の非調和成分については、その調和成分の振幅スペクトルに基づいて非調和成分の振幅スペクトルを計算し、それを前記平坦なスペクトルに乗ずることにより、非調和成分の振幅スペクトルを得ることを特徴とする請求項９記載の歌唱合成装置。

【請求項１１】　前記音韻データベース中の一部の伸ばし音についての音声素片については、その非調和成分の振幅スペクトルを記憶せず、他の伸ばし音の音声素片に記憶されている前記平坦なスペクトルを使用して、その伸ばし音を合成することを特徴とする請求項９あるいは１０に記載の歌唱合成装置。

【請求項１２】　前記調和成分の振幅スペクトルに基づいて非調和成分の振幅スペクトルを計算するときに、ハスキー度を制御するパラメータに応じて前記計算する非調和成分の振幅スペクトルの０Hzにおけるゲインを制御することを特徴とする請求項１０記載の歌唱合成装置。

【請求項１３】　歌唱音合成時に、伸ばし音の非調和成分の振幅スペクトルに、その伸ばし音区間内における代表振幅スペクトルの逆数を乗算して平坦なスペクトルを作成し、その伸ばし音の調和成分の振幅スペクトルに基づいてハスキー度を制御するパラメータに応じた振幅スペクトルを計算し、該振幅スペクトルと前記作成した平坦なスペクトルとを乗ずることにより得られた振幅スペクトルをその伸ばし音の非調和成分の振幅スペクトルとして使用することを特徴とする請求項５記載の歌唱合成装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、歌声を合成する歌唱合成装置に関する。

【０００２】

【従来の技術】従来より、歌声を合成しようとする試みは幅広く行われてきた。そのうちの１つは、規則音声合成の応用で、音符の音程に対応する音高データと歌詞データを入力とし、テキスト音声合成用の規則音声合成器を用いて合成するものである。多くの場合、音素（あるいは音韻：phoneme）あるいは２つ以上の音素を含む音素連鎖を単位とする生波形データあるいはそれを分析しパラメータ化したものをデータベースに蓄積し、合成時に必要な音声素片（音素あるいは音素連鎖）を選択し、接続、合成するものである。例えば、特開昭６２－６２

９９号公報、特開平１０－１２４０８２号公報、特開平１１－１１８４４９０号公報などを参照されたい。しかしながら、これらの技術は、本来、話し言葉を合成することを目的としているため、歌声を合成する場合には品質が必ずしも満足することのできるものではなかった。

【０００３】例えば、ＰＳＯＬＡ（Pitch-Synchronous OverLap and Add）に代表される波形重畳合成方式では、合成歌唱音の了解度は良好であるが、歌唱音の品質を最も左右する音を伸ばしている部分が不自然になってしまう場合が多い、歌唱音声に必要不可欠なビブラートやピッチの微妙な変動を行なった場合に不自然な合成音になってしまうことが多いという問題点があった。また、大規模コーパスベースの波形接続型音声合成器を使って歌唱音声を合成しようとすれば、もとの波形を原則として全く加工せずに接続して出力するため、天文学的数字の素片データが必要となる。

【０００４】一方、はじめから歌声の合成を目的とした合成器も考案されている。例えば、フォルマント合成方式による合成方式が知られている（特開平３－２００３００号公報）。これは、伸ばし音の品質やビブラートやピッチ変化の自由度は大きいが、合成音（特に子音部分）の明瞭度が低く、品質は必ずしも満足できるものではない。

【０００５】ところで、米国特許第５０２９５０９号明細書に示されるように、オリジナルの音を２つの成分、すなわち調和成分（deterministic component）と非調和成分（stochastic component）で表わすモデルを使用して楽音の分析および合成を行なう、スペクトルモデリング合成（ＳＭＳ：Spectral Modeling Synthesis）と呼ばれる技術が知られている。このＳＭＳ分析合成によれば、楽音の音楽的特徴を良好に制御することができると同時に、歌声の場合には、非調和成分の利用により、子音部分でも高い明瞭度が得られることが期待できる。したがって、この技術を歌声の合成に応用すれば、高い明瞭度と音楽性を併せ持った合成音が得られることが期待される。現に、特許第２９０６９７０号では、ＳＭＳ分析合成技術に基づき音を合成する手法についての具体的応用の提案が行われているが、同時にＳＭＳ技術を歌唱合成（シンギング・シンセサイザ）に利用する場合の方法論についても述べられている。

【０００６】前記特許第２９０６９７０号に提案されている手法を適用した歌唱合成装置について、図１７を参照して説明する。図１７において、音韻データベース１００は、入力音声をＳＭＳ分析および区間切り出し部１０３において、ＳＭＳ分析し、音声素片（音素あるいは音素連鎖）ごとに切り出して、記憶することにより作成される。データベース１００中の音声素片データ（音素データ１０１、音素連鎖データ１０２）は、時系列に並べられた単一あるいは複数のフレーム列のデータから構成され、各フレームに対応するＳＭＳデータ、すなわち、調和成分のスペクトル包絡、非調和成分のスペクトル包絡と位相スペクトルなどの時間的変化が記憶されている。歌唱音を合成するときには、所望の歌詞を構成する音素列を求め、音素→素片変換部１０４により、その音素列を構成するのに必要な音声素片（音素あるいは音素連鎖）を決定し、前記データベース１００から必要な音声素片のＳＭＳデータ（調和成分と非調和成分）を読み出す。そして、素片接続部１０５において読み出した音声素片のＳＭＳデータを時系列的に接続し、調和成分については調和成分生成部１０６においてその楽曲のメロディーに対応するピッチ情報に応じて、そのスペクトル包絡の形状を保ったまま、所望のピッチを有する倍音成分を生成する。例えば、「サイタ」（saita）と合成する場合には、[#s]，[s]，[s-a]，[a]，[a-i]，[i]，[i-t]，[t]，[t-a]，[a]，[a#]という素片を接続し、素片の接続により得られたＳＭＳデータに含まれるスペクトル包絡の形状を保ったまま、所望のピッチの調和成分を生成する。そして、この生成した調和成分と非調和成分とを合成手段１０７で加算し、時間領域のデータに変換することにより、合成音声を得る。

【０００７】

【発明が解決しようとする課題】このようにＳＭＳ技術を利用することにより、了解度が良好で、かつ、伸ばしている部分についても自然な合成歌唱音を得ることが可能となる。しかし、上記特許第２９０６９７０号で述べられている方式は、あまりにも原始的かつ単純であり、その方式のまま歌声を合成すると、次のような問題点が生じる。

・有声音の調和成分のスペクトル包絡の形状がピッチによって若干変化するため、分析時とは異なるピッチで合成する場合に、そのままでは良い音色が得られない。

・ＳＭＳ分析を行う場合、有声音の場合に調和成分を取り去っても残差成分にわずかながら調和成分が残るため、上記のように同じ残差成分（非調和成分）をそのまま用いて元の音とは異なるピッチの歌唱音で合成すると残差成分が浮いて聴こえたり、ノイズに聴こえる原因となる。

・ＳＭＳの分析結果としての音素データ、音素連鎖データをそのまま時間的に重ね合わせているため、音を伸ばす時間や音素間の移り変わりの時間の調整ができない。すなわち、所望のテンポで歌わせることができない。

・音素あるいは音素連鎖の接続時に雑音が発生しやすい。

【０００８】そこで本発明は、上記特許第２９０６９７０号において提案されているＳＭＳ技術を歌唱合成に利用する場合の手法を具体化し、さらに合成音の品質について大幅な改良を加え、上述の各問題点を解決した歌唱合成装置を提供することを目的としている。また、前記データベースのサイズを小さくすることができるとともに、データベース作成の効率を向上させた歌唱合成装置

を提供することを目的としている。さらに、合成音声のハスキーさの度合いを調整することのできる歌唱合成装置を提供することを目的としている。

【０００９】

【課題を解決するための手段】上記目的を達成するために、本発明の歌唱合成装置は、音素あるいは２つ以上の音素のつながりである音素連鎖である音声素片について調和成分のデータと非調和成分のデータを記憶した音韻データベースを有し、歌詞に対応した音声素片データを前記音韻データベースから読み出して接続することにより、歌唱音を合成する歌唱合成装置であって、目的のテンポや歌い方に合うように前記音韻データベースから読み出した音声素片データの時間長を調整する継続時間調整手段と、目的のピッチに合うように前記音韻データベースから読み出した音声素片データの前記調和成分および前記非調和成分を調整する調整手段とを有するものである。また、前記音声素片データを接続するときに、調和成分、非調和成分それぞれについてスムージング処理あるいはレベル調整処理を行なう素片レベル調整手段を有するものである。さらに、前記音韻データベース中には、同一の音素または音素連鎖について、ピッチ、ダイナミクス、テンポの異なる複数の音声素片データが記憶されているものである。さらにまた、前記音韻データベース中には、母音などの伸ばし音からなる音声素片データ、子音から母音あるいは母音から子音への音素連鎖からなる音声素片データ、子音から子音への音素連鎖からなる音声素片データおよび母音から母音への音素連鎖からなる音声素片データが記憶されているものである。

【００１０】さらにまた、前記調和成分のデータと前記非調和成分のデータは、その素片の区間に含まれるフレーム列の各フレームに対応する周波数領域のデータ列として記憶されているものである。さらにまた、前記継続時間調整手段は、音声素片に含まれるフレーム列中の１または複数のフレームを繰り返すこと、あるいは、フレームを間引くことにより所望の時間長のフレーム列を生成するものである。さらにまた、前記継続時間調整手段は、非調和成分のフレームを繰り返すときに、合成時に時間的に逆行する場合には、その非調和成分の位相スペクトルの位相を反転させるものである。さらにまた、歌唱音合成時に、調和成分について、音声素片データに含まれている調和成分のスペクトル包絡の概形を保ったままピッチだけを所望のピッチに変換する調和成分生成手段を有するものである。

【００１１】さらにまた、前記音韻データベース中に記憶される音声素片データのうち伸ばし音に対応する音声素片については、非調和成分の振幅スペクトルとして、その非調和成分の振幅スペクトルにその伸ばし音の区間を代表するスペクトルの逆数を乗算することにより得られた平坦なスペクトルを記憶しているものである。さらにまた、歌唱音合成時に、伸ばし音の非調和成分については、その調和成分の振幅スペクトルに基づいて非調和成分の振幅スペクトルを計算し、それを前記平坦なスペクトルに乗ずることにより、非調和成分の振幅スペクトルを得るものである。さらにまた、前記音韻データベース中の一部の伸ばし音についての音声素片については、その非調和成分の振幅スペクトルを記憶せず、他の伸ばし音の音声素片に記憶されている前記平坦なスペクトルを使用して、その伸ばし音を合成するものである。さらにまた、前記調和成分の振幅スペクトルに基づいて非調和成分の振幅スペクトルを計算するときに、ハスキー度を制御するパラメータに応じて前記計算する非調和成分の振幅スペクトルの０Ｈｚにおけるゲインを制御するものである。

【００１２】さらにまた、歌唱音合成時に、伸ばし音の非調和成分の振幅スペクトルに、その伸ばし音区間内における代表振幅スペクトルの逆数を乗算して平坦なスペクトルを作成し、その伸ばし音の調和成分の振幅スペクトルに基づいてハスキー度を制御するパラメータに応じた振幅スペクトルを計算し、該振幅スペクトルと前記作成した平坦なスペクトルとを乗ずることにより得られた振幅スペクトルをその伸ばし音の非調和成分の振幅スペクトルとして使用するものである。

【００１３】

【発明の実施の形態】本発明の歌唱合成装置は、入力音声をＳＭＳ分析し、調和成分、非調和成分のＳＭＳデータを求め、必要な区間を切り出して音素ごと、および音素連鎖ごとにまとめた音韻データベースを持つ。このデータベース内には、見出しとして音素または音素連鎖の情報に加え、その音声素片のピッチを示す情報、および、ダイナミクスやテンポなどの音楽表現を示す情報も含まれる。ここで、ダイナミクス情報は、その音声素片（音素または音素連鎖）がフォルテの音であるのかメゾフォルテの音であるのかといった感覚的な情報であってもよいし、あるいは、その素片のレベルを示す物理的な情報であってもよい。また、前記データベース作成のために、入力歌唱音声を非調和成分、調和成分に分解して分析するＳＭＳ分析手段を備える。また、必要とする音素または音素連鎖（素片）を切り出すための手段（自動、手動を問わない）を備える。

【００１４】図１を参照して、前記音韻データベース作成の例について説明する。図１において、１０は音韻データベースであり、前述した音韻データベース１００と同様に、入力歌唱音声をＳＭＳ分析部１３でＳＭＳ分析し、区間切り出し部１４により音素または音素連鎖（音声素片）毎に切り出された各素片毎のＳＭＳデータ（その素片に含まれている各フレームのＳＭＳデータ）が格納されている。ただし、この音韻データベース１０においては、素片データが異なるピッチ、異なるダイナミクス、異なるテンポ毎に別個のデータとして記憶されている。

【００１５】なお、日本語の歌詞を歌唱させる場合には、音声素片は例えば母音だけのデータ（１フレームあるいは複数のフレーム）と、子音から母音へのデータ（複数フレーム）あるいは母音から子音へのデータ（複数フレーム）と、子音から子音へのデータ（複数フレーム）と、母音から母音へのデータ（複数フレーム）とからなる。規則合成などの音声合成装置においては、通常、音節よりも長いＶＣＶ（母音・子音・母音）あるいはＣＶＣ（子音・母音・子音）などを音韻データベースに記録する単位としているが、特に歌唱音の合成を目的としている本発明の歌唱合成装置においては、歌唱においてよく現れる母音などを長く発音する伸ばし音のデータ、子音から母音（ＣＶ）あるいは母音から子音（ＶＣ）のデータ、子音から子音のデータ、および、母音から母音のデータを音韻データベースに格納している。

【００１６】前記ＳＭＳ分析部１３は、オリジナルの入力歌唱音声をＳＭＳ分析し、各フレーム毎のＳＭＳ分析データを出力する。すなわち、入力音声を一連の時間フレームに分け、各フレーム毎にＦＦＴなどにより周波数分析する。その結果得られた周波数スペクトル（複素スペクトル）から振幅スペクトルと位相スペクトルを求め、振幅スペクトルのピークに対応する特定の周波数のスペクトルを線スペクトルとして抽出する。このとき、基本周波数およびその整数倍の周波数の近傍の周波数を持つスペクトルを線スペクトルとする。この抽出した線スペクトルが前記調和成分に対応している。そして、上記のようにして抽出した線スペクトルをそのフレームの入力波形のスペクトルから減算することにより、残差スペクトルを得る。あるいは、前記抽出した線スペクトルから合成した調和成分の時間波形データをそのフレームの入力波形データから減算して残差成分の時間波形データを得、これを周波数分析することにより残差スペクトルを得る。このようにして得た残差スペクトルが、前記非調和成分（ストカスティック成分）に対応する。

【００１７】なお、前記ＳＭＳ分析に用いるフレーム周期は、一定の固定長であってもよいし、あるいは、入力音声のピッチ等に応じてその周期を変更する可変長の周期であっても良い。フレーム周期を可変長とする場合には、固定長の第１のフレーム周期で入力音声を処理してそのピッチを検出し、その結果に応じたフレーム周期で入力音声を再処理する、あるいは、そのフレームの前のフレームの分析結果から得たピッチにより後続するフレームの周期を変更するなどの手法を採用すればよい。

【００１８】前記ＳＭＳ分析部１３から各フレーム毎に出力されるＳＭＳ分析データは、区間切り出し部１４において、音韻データベースに記憶する音声素片の長さに対応するように切り出される。すなわち、歌唱音の合成に最も適するように、母音の音素、母音と子音あるいは子音と母音の音素連鎖、子音と子音の音素連鎖、および、母音と母音の音素連鎖が手動あるいは自動的に切り出される。ここで、母音の音素として、その母音を伸ばして歌唱している長区間のデータ（伸ばし音）も切り出される。また、この区間切り出し部１４において、前記ＳＭＳ分析結果からその入力音声のピッチを検出する。このピッチ検出は、その素片に含まれるフレームの調和成分のうちの低次の線スペクトルの周波数から平均ピッチを求め、これを全フレームについて平均することにより行なわれる。

【００１９】このようにして、各素片ごとにその調和成分のデータおよび非調和成分のデータを切り出し、さらに、その入力歌唱音声のピッチ、音楽表現を表わすダイナミクス、テンポなどの情報を見出しとして付加して前記音韻データベース１０に格納する。図１には、このようにして作成された音韻データベース１０の一例を示しており、音韻データベース１０中に音素に対応する音素データ領域１１および音素連鎖に対応する音素連鎖データ領域１２が示されている。そして、前記音素データ領域１１には、母音[ａ]の伸ばし音に対してピッチ周波数130Hz，150Hz，200Hz，220Hzの４通りの音素データ、母音[ｉ]の伸ばし音に対してピッチ周波数140Hz，180Hz，300Hzの３通りの音素データが格納されている様子が示されている。また、前記音素連鎖データ領域１２には、音素[ａ]と[ｉ]のつながりを示す音素連鎖[ａ-ｉ]に対してピッチ周波数130Hzと150Hzの２通り、音素連鎖[ａ-ｐ]に対して120Hzと220Hzの２通り、音素連鎖[ａ-ｓ]に対して140Hzと180Hz、音素連鎖[ａ-ｚ]に対して100Hzの各音素連鎖データが格納されている様子が示されている。なお、ここでは、同一の音素あるいは音素連鎖に対してピッチが異なるデータを格納している場合を示しているが、前述のように、その入力歌唱音声のダイナミクスやテンポなどの音楽表現が異なるデータについても、同様に、異なるデータとして記憶する。

【００２０】なお、それぞれの素片データに含まれている調和成分と非調和成分を表わすデータは、前記区間切り出し部１４により各素片ごとに切り出された前記ＳＭＳ分析部１３からのＳＭＳデータ、すなわち、調和成分については、その素片に含まれる各フレームの全てのスペクトル包絡（線スペクトル（倍音系列）の強度（振幅）および位相のスペクトル）をそのまま記憶する、あるいは、スペクトル包絡そのものではなく、スペクトル包絡を何らかの関数で表現したものとして記憶する、のいずれの方法で記憶しても良い。あるいは、調和成分を逆変換した時間波形の形で記憶しても良い。また、非調和成分についても、その素片に対応する区間の各フレームの強度スペクトル（振幅スペクトル）および位相スペクトルとして記憶しても良いし、その区間の時間波形データそのものの形で記憶しても良い。また、上記各記憶形式は固定である必要はなく、素片毎に、あるいは、その区間の音声の性質（例えば、鼻音、摩擦音、破裂音など）に応じてその記憶形式を異ならしめるようにしても

よい。なお、以下の説明では、前記調和成分のデータはスペクトル包絡の形式で記憶し、非調和成分はその振幅スペクトルおよび位相スペクトルの形式で記憶しているものとして説明する。このような記憶形式の場合には、必要とされる記憶容量を少なくすることができる。このように、本発明の歌唱合成装置における音韻データベース１０には、同一の音素あるいは音韻に対して異なるピッチあるいはダイナミクス、テンポなどの音楽表現に対応する複数のデータが格納されている。

【００２１】次に、このように作成された音韻データベース１０を用いた歌唱音の合成処理について図２を参照して説明する。図２において、１０は前述した音韻データベースである。また、２１は音素→素片変換手段であり、歌唱音を合成すべき楽曲の歌詞データに対応する音素列を、前記音韻データベース１０を検索するための素片に変換するものである。例えば、「s_a_i_t_a」という音素列の入力に対し、素片列[s] [s-a] [a] [a-i] [i] [i-t] [t] [t-a] [a]を出力する。２２は、前記楽曲のメロディデータなどに含まれているピッチやダイナミクスやテンポなどのコントロールパラメータに基づいて、前記音韻データベース１０から読み出された素片データのうちの調和成分のデータの調整を行う調和成分調整手段、２３は前記非調和成分のデータに対して調整を行う非調和成分調整手段である。２４は、前記調和成分調整手段２２および前記非調和成分調整手段２３からの素片データの継続時間を変更する継続時間調整手段、２５は前記継続時間調整手段２４からの各素片データのレベルの調整を行う素片レベル調整手段、２６は前記素片レベル調整手段２５によりレベル調整された各素片データを時系列に接続する素片接続手段、２７は前記素片接続手段２６により接続された素片データのうちの調和成分のデータ（スペクトル包絡情報）に基づいて所望のピッチの調和成分（倍音成分）を生成する調和成分生成手段、２８は前記調和成分生成手段２７で生成された倍音成分と前記素片接続手段２６から出力される非調和成分とを合成する加算手段である。この加算手段２８の出力を時間領域の信号に変換することにより、合成音声が得られる。

【００２２】以下、上記各ブロックにおける処理について詳細に説明する。前記音素→素片変換手段２１は、入力歌詞をもとに変換した音素列から素片列を生成し、それにより、音韻データベース１０中の音声素片（音素や音素連鎖）の選択を行なう。前述のように、同じ音素や音素連鎖であっても、ピッチ、ダイナミクス、テンポなどに対応してデータベース中に複数のもの（音声素片データ）が格納されており、素片選択時に各種コントロールパラメータに応じて最適なものを選択する。また、選択するのではなくいくつかの候補を選択し、それらの補間により合成に用いるＳＭＳデータを求めるようにしても良い。選択された音声素片にはＳＭＳ分析の結果としての調和成分と非調和成分が格納されている。この内容は、ＳＭＳデータ、すなわち、調和成分のスペクトル包絡（強度と位相）と非調和成分のスペクトル包絡（強度と位相）または波形そのものが入っている。これらの内容を元に、所望のピッチ、要求される継続時間に合うように調和成分、非調和成分を生成する。例えば、所望のピッチに合うように調和・非調和成分のスペクトル包絡を補間などにより求めたり、スペクトル形状を変形させる。

【００２３】［調和成分の調整］前記調和成分調整手段２２では、調和成分の調整処理を行う。有声音の場合、調和成分については、ＳＭＳ分析結果である調和成分の強度および位相のスペクトル包絡が入っている。素片が複数の場合は、その中から所望のコントロールパラメータ（ピッチなど）に最適なものを選択するか、あるいは複数の素片の中から補間などの操作により所望のコントロールパラメータに適したスペクトル包絡を求める。また、得られたスペクトル包絡をさらに別のコントロールパラメータに対応して何らかの方法で変形させても良い。また、耳障りとなる音を軽減させたり、音に特徴を持たせたりするため、一定の帯域のみ通過させるようなフィルターをかけても良い。なお、無声音の場合は調和成分はない。

【００２４】［非調和成分の調整］有声音のＳＭＳ分析結果の非調和成分には、元のピッチの影響が残っているので、別のピッチの音を合成する場合には、音が不自然になってしまう場合がある。これを防ぐために、非調和成分の低域成分に対し、所望のピッチに合うような操作を行なう必要がある。前記非調和成分調整手段２３では、この操作を行う。図３を参照して、この非調和成分に対する調整操作について説明する。図３の（ａ）は、有声音をＳＭＳ分析したときに得られる非調和成分の振幅スペクトルの例である。この図に示すように、調和成分の影響を完全に取り去ることは難しく、倍音付近に若干の山ができている。この非調和成分をそのまま用いて、もとのピッチとは別のピッチで音声を合成すると、低域の倍音付近の山々が知覚され、調和成分とうまく溶け合わずに耳障りな音に聴こえる場合がある。そこで、非調和成分の周波数をピッチの変化に合わせて変えてやればよいが、高域の非調和成分はもともと調和成分の影響が少ないので、もともとの振幅スペクトルをそのまま用いることが望ましい。つまり、低域においては求めるピッチにしたがって周波数軸の圧縮・伸長を行なえばよい。ただし、このときに元の音色は変化させてはならない。つまり、振幅スペクトルの概形を保ったままこの処理を行なう必要がある。

【００２５】図３の（ｂ）は、上述の処理を行なった結果を示す図である。この図に示すように、低域の３つの山は所望のピッチに従い、右に移動されている。中域の山の間隔は狭められ、高域の山はそのままとなってい

る。それぞれの山は、破線で示す振幅スペクトルの概形を保つように高さが調整される。なお、無声音の場合は、元のピッチの影響はないので、上記の操作は必要ない。また、得られた非調和成分に対し、コントロールパラメータに対応してさらに何らかの操作（例えば、スペクトル包絡形状の変形など）を行なってもよい。また、耳障りとなる音を軽減させたり、音に特徴を持たせたりするため、一定の帯域のみ通過させるようなフィルターをかけてもよい。

【００２６】［継続時間調整］さて、このままでは素片の持つもともとの長さをそのまま使うことになるため、一定のタイミングでしか歌声を合成することができない。そこで、求めるタイミングに応じて必要ならば素片の継続長を変更する必要がある。例えば、音素連鎖の場合には、素片内に含まれるフレームを間引くことで素片の長さは短くなり、重複させることで長くすることができる。また、例えば、音素が１つの場合（伸ばし音の場合）には、素片内のフレーム一部だけを用いれば伸ばし部分は短くなり、素片内を繰り返すことで長くすることができる。

【００２７】伸ばし音の場合に素片内を繰り返すとき、単に一方向だけ繰り返すよりも一方向に進んで逆方向に戻り、再び元の方向に進む（すなわち、一定区間あるいはランダムな区間内をループする）ということを繰り返すほうが、つなぎ目の雑音が軽減できることが知られているが、非調和成分がフレーム（固定あるいは可変長）ごとに区切られて周波数領域で記憶されている場合には、周波数領域のフレームデータをそのままの形で繰り返して波形を合成するのは問題である。これは、時間的に逆方向に進むときにはフレーム内の波形自体も時間的に逆になるようにしなければならないからである。時間的に逆方向に進む波形を元の周波数領域のフレームデータから生成するには、周波数領域の位相を反転させて時間領域に変換すればよい。図４は、この様子を示す図である。

【００２８】図４の（ａ）は、もともとの非調和成分の波形を示す図である。図に示す繰り返し区間t1からt2まで進み、t2に達した後は時間的に逆方向に進み、再びt1に達した後は順方向に進む、ということを繰り返して伸ばし音のための非調和成分を生成するものとする。非調和成分は、前述のように、固定あるいは可変長のフレームごとに区切られて周波数成分で記憶されている。時間領域の波形を生成するには、周波数領域のフレームデータを逆ＦＦＴし、窓関数を掛けてオーバーラップさせながら合成すればよい。ここで、時間的に逆方向にフレームを読み込んで合成する場合、周波数領域のフレームデータをそのまま時間領域に変換すると、図４（ｂ）に示すように、フレーム内の波形は時間的に元のままフレームの順番だけが逆になった波形になってしまい、不連続となって雑音や歪みなどの原因となる。

【００２９】これを解決するためには、フレームデータから時間領域の波形を求める際に、時間的に逆の波形が生成されるようにあらかじめフレームデータを加工すればよい。もとの波形をｆ(t)（便宜上、無限に続く波形と考える）、時間的に逆方向になる波形をｇ(t)とし、それぞれのフーリエ変換をＦ(ω)、Ｇ(ω)とすると、ｇ(t)＝ｆ(-t)であり、かつ、ｆ(t)、ｇ(t)ともに実関数なので、

Ｇ(ω)＝Ｆ(ω)*（*は複素共役を示す）

が成立する。振幅と位相で表わした場合に、複素共役は位相を逆にしたものになるので、時間的に逆の波形を生成するためには、周波数領域のフレームデータの位相スペクトルをすべて逆にすれば良いことがわかる。このようにすれば、図４の（ｃ）に示すように、フレーム内部も時間的に逆の波形となり、雑音や歪みが生じない。

【００３０】前記継続時間調整手段２４では、上述のような素片の圧縮処理（フレームの間引き）、伸長処理（フレームの繰り返し）およびループ処理（伸ばし音の場合）を行なう。これにより、読み出した各素片の継続時間（すなわちフレーム列の長さ）を所望の長さに調整することができる。

【００３１】［素片レベル調整］さらに、素片と素片の接続部分で調和・非調和の各成分のスペクトル包絡の形状に差がありすぎる場合は、雑音として聴こえる怖れがある。複数のフレームをかけて接続部分をスムージングすることによりこれを解消することができる。このスムージング処理について図５～図７を参照して説明する。非調和成分については、素片の接続部に音色やレベルのばらつきがあっても、比較的聴こえにくいため、ここでは、調和成分のみスムージングするものとする。このとき、データを扱いやすくして計算を簡単にするために、調和成分のスペクトル包絡を図５に示すように、直線あるいは指数関数で表現した傾き成分と指数関数などで表現した共鳴成分とに分けて考えることとする。ここで、共鳴成分の強度は傾き成分を基準に計算するものとし、傾き成分と共鳴成分を足し合わせてスペクトル包絡を表わすものとする。すなわち、調和成分を前記傾き成分と共鳴成分とを用いたスペクトル包絡を表わす関数で表現している。ここで、前記傾き成分を０Hzまで延長した値を傾き成分のゲインと称することとする。

【００３２】このとき、図６に示すような２つの素片[a-i]と[i-a]とを接続するものとする。各素片は、もともと別の録音から採集したものであるため、接続部のｉの音色とレベルにミスマッチがあるため、図６に示すように、接続部分で波形の段差が発生し、ノイズとして聴こえる。そこで、その接続部を中心とし前後に何フレームかかけて、それぞれの素片に含まれる傾き成分と共鳴成分の各パラメータをクロスフェードしてやれば、接続部分での段差が消え去り、ノイズの発生を防止することができる。各パラメータをクロスフェードするため

には、図７に示すように、接続部分で0.5となるような関数（クロスフェードパラメータ）を両素片の各パラメータに掛けて足し合わせてやればよい。図７に示す例では、第１の共鳴成分の（傾き成分を基準とした）強度の[a-i]，[i-a]の各素片における動きと、これをクロスフェードする例を示している。このように、各パラメータ（この場合は、各共鳴成分）にクロスフェードパラメータを乗算して足し合わせることにより素片の接続部におけるノイズの発生を防止することができる。

【００３３】また、上記のようにクロスフェードする代わりに、素片の接続部分で前後の振幅がほぼ同じになるように、素片の調和・非調和の各成分のレベル調整を行っても良い。レベル調整は、素片の振幅に対し、一定あるいは時変の係数を掛けることにより行なうことができる。上記と同様に、[a-i]と[i-a]を接続して合成する場合を例にとって、レベル調整の一例につき説明する。ここでは、前記各素片の傾き成分のゲインを合わせることを考える。図８の（ａ）、（ｂ）に示すように、まず、[a-i]と[i-a]の各素片について、その最初のフレームと最終フレームの間の傾き成分のゲインを直線補間したもの（図中の破線）を基準に、実際の傾き成分のゲインとの差分を求める。次に、[a]，[i]の各音韻の代表的なサンプル（傾き成分および共鳴成分の各パラメータ）を求める。これは、例えば、[a-i]の最初のフレームと最終フレームのデータを用いても良い。この代表サンプルをもとに、まず、パラメータを直線補間したものを求め、次いで、上で求めた差分を足し込んでいけば、図８の（ｃ）に示すように、境界ではかならず全てのパラメータが同じになるため、傾き成分のゲインの不連続は発生しない。共鳴成分のパラメータなど他のパラメータについても、同様に不連続を防止することができる。なお、以上に述べた方法によらず、例えば、調和成分のデータを波形データに変換し、時間領域でレベル調整などを行うようにしてもよい。

【００３４】前記素片レベル調整手段２６において、上述した素片間のスムージングあるいはレベル調整処理が行われた後、素片接続手段２６で素片接続処理が行なわれる。そして、調和成分生成手段２７において、得られた調和成分スペクトル包絡を保ったまま所望のピッチに対応する倍音列を発生することにより、実際の調和成分が得られ、それに非調和成分を足し合わせることにより、合成歌唱音が得られる。そして、これを時間領域の信号に変換する。例えば、調和・非調和の両成分を周波数成分で持っている場合には、両成分を周波数領域で足し合わせ逆ＦＦＴと窓掛けおよびオーバーラップを行なうことにより、合成波形が得られる。なお、両成分を別々に逆ＦＦＴを窓掛けおよびオーバラップを行い、後で足し合わせてもよい。また、調和成分については、各倍音に対応する正弦波を生成し、逆ＦＦＴと窓掛けおよびオーバーラップにより求められた非調和成分と足しあわせても良い。

【００３５】図９は、前記図２に示した本発明の歌唱合成装置の一実施の形態の構成をより詳細に示す機能ブロック図である。この図において、前記図２と同一の構成要素には同一の符号を付す。また、この例では、音韻（音声素片）データベース１０中には、調和成分はフレーム毎の振幅スペクトル包絡情報、非調和成分はフレーム毎の振幅スペクトル包絡情報と位相スペクトル包絡情報が含まれているものとする。図９において、３１は、歌声を合成すべき楽曲の楽譜データから歌詞データとメロディデータを分離する歌詞・メロディー分離手段、３２は前記歌詞・メロディー分離手段３１からの歌詞データを音声記号（音素）列に変換する歌詞音声記号変換手段であり、この歌詞音声記号変換手段３２からの音素列は前記音素（音声記号）素片変換手段２１に入力される。また、演奏を制御するテンポなどの各種コントロールパラメータが入力可能とされており、前記歌詞・メロディー分離手段３１で楽譜データから分離されたピッチ情報と強弱記号などのダイナミクス情報および前記コントロールパラメータはピッチ決定手段３３に入力され、歌唱音のピッチやダイナミクスおよびテンポが決定される。前記音素素片変換手段２１からの素片情報および前記ピッチ決定手段からのピッチ、ダイナミクス、テンポなどの情報は、素片選択手段３４に供給され、該素片選択手段３４は、前記音声素片データベース（音韻データベース）１０から最も適切な素片データを検索して出力する。このとき、検索条件に完全に一致する素片データが記憶されていないときには、類似する１または複数の素片データを読み出す。

【００３６】前記素片選択手段３４から出力された素片データの内の調和成分のデータは、調和成分調整手段２２に供給される。前記素片選択手段３４により読み出された素片データが複数の場合には、この調和成分調整手段２２におけるスペクトル包絡補間部３５で前記検索条件に合致するように補間処理を行ない、さらに、必要に応じて、スペクトル包絡変形部３６で前記コントロールパラメータに対応してスペクトル包絡の形状を変形する。一方、前記素片選択手段３４から出力された素片データのうちの非調和成分のデータは非調和成分調整手段２３に入力される。この非調和成分調整手段２３には、前記ピッチ決定手段３３からのピッチ情報が入力されており、前記図３に関して説明したように、非調和成分の低域成分に対してピッチに応じた周波数軸の圧縮あるいは伸長処理を行なう。すなわち、バンドパスフィルター３７により、非調和成分の振幅スペクトルおよび位相スペクトルを低域、中域、高域に３分割し、低域および中域については周波数軸圧縮・伸長部３８および３９でそれぞれピッチに対応した周波数軸の圧縮あるいは伸長を行なう。この周波数軸の圧縮あるいは伸長処理が行なわれた低域および中域の信号およびこのような操作がなさ

れない高域の信号は、ピーク調整部４０に供給され、この非調和成分のスペクトル包絡の形状を維持するように、そのピーク値が調整される。

【００３７】前記調和成分調整手段２２からの調和成分データおよび前記非調和成分調整手段２３からの非調和成分データは、継続時間長調整手段２４に入力される。そして、この継続時間長調整手段２４において、前記メロディー情報および前記テンポ情報により決定される発音時間長に応じて素片の時間長の変更が行なわれる。前述のように、素片データの継続時間を短くする場合には、時間軸圧縮・伸長部４３でフレームの間引きを行い、継続時間を長くするときには、ループ部４２で、前記図４に関して説明したループ処理を行なう。前記継続時間長調整手段２４で継続時間長を調整された素片データは、レベル調整手段２５で前記図５～図８に関して説明したようなレベル調整処理を施され、素片接続手段２６で調和成分、非調和成分それぞれ時系列に接続される。

【００３８】前記素片接続手段２６で接続された素片データの調和成分（スペクトル包絡情報）は調和成分生成手段２７に入力される。この調和成分生成手段２７には、前記ピッチ決定手段３３からのピッチ情報が供給されており、前記スペクトル包絡情報に従った前記ピッチ情報に対応する倍音成分を生成する。これにより、そのフレームの実際の調和成分が得られる。そして、前記素片接続手段２６からの非調和成分の振幅スペクトル包絡情報および位相スペクトル包絡情報と、前記調和成分生成手段２７からの調和成分の振幅スペクトルを加算器２８で合成する。そして、このように合成された各フレームに対応する周波数領域の信号を逆フーリエ変換手段（逆ＦＦＴ手段）５１で時間領域の波形信号に変換し、さらに、窓掛け手段５２でフレーム長に対応した窓関数を乗算し、さらに、オーバーラップ手段５３により各フレーム毎の波形信号をオーバーラップさせながら合成する。そして、このように合成した時間波形信号をＤ／Ａ変換手段５４でアナログ信号に変換し、増幅器５５を介してスピーカ５６から出力する。

【００３９】さらに、図１０は、前記図９に示した具体例を動作させるためのハードウエア装置の一例を示す図である。この図において、６１はこの歌唱合成装置全体の動作を制御する中央処理装置（ＣＰＵ）、６２は各種プログラムや定数などが記憶されているＲＯＭ、６３はワークエリアや各種データを記憶するＲＡＭ、６４はデータメモリ、６５は所定のタイマ割込みなどを発生させるタイマ、６６は前記演奏すべき楽曲の楽譜データや歌詞データなどを入力する歌詞・メロディー入力部、６７は演奏に関する各コントロールパラメータなどを入力するコントロールパラメータ入力部、６８は各種情報を表示する表示部、６９は前記合成された歌唱データをアナログ信号に変換するＤ／Ａ変換器、７０は増幅器、７１はスピーカ、７２は前記各構成要素間を接続するバスである。ここで、前記ＲＯＭ６２あるいはＲＡＭ６３上に前記音韻データベース１０がロードされ、歌詞・メロディ入力部６６およびコントロールパラメータ入力部６７から入力されたデータに従い、前述のように歌唱音の合成を行ない、合成音はスピーカ７１から出力される。この図１０に示す構成は、通常の汎用コンピュータと同一の構成とされており、本発明の歌唱合成装置の上記各機能部は、汎用コンピュータのアプリケーションプログラムとしても実現することができる。

【００４０】さて、上述した実施の形態においては、前記音韻データベース１０に格納されている素片データは、ＳＭＳデータ、代表的な例では、調和成分の単位時間（フレーム）毎のスペクトル包絡、および、非調和成分のフレーム毎の振幅スペクトルおよび位相スペクトルであった。そして、前述のように、母音などの伸ばし音の素片データを記憶することにより、高品質の歌唱音を合成することができるものであった。しかしながら、特に伸ばし音の場合には、その伸ばし音の区間全ての時刻（フレーム）における調和成分および非調和成分が記憶されているため、データ量が大きくなってしまうという問題がある。調和成分の場合は、基本ピッチの整数倍の周波数ごとにデータを持てばよいので、例えば基本ピッチが150Hz、最大周波数が22025zとして、１５０の周波数についての振幅データ（あるいは位相も）を持つ必要がある。これに対し、非調和成分の場合にはさらに多くのデータが必要で、振幅スペクトル包絡と位相スペクトル包絡を全ての周波数について持つ必要がある。１フレーム内のサンプリング点数を1024点とした場合、1024の周波数について振幅および位相のデータが必要となる。特に、伸ばし音については、伸ばし音区間中の全てのフレームについてデータを持つ必要があるため、データの大きさは非常に大きなものとなってしまう。また、伸ばし音の区間のデータは各音素ごとに用意する必要があるのに加え、上述のように、自然性を上げるためにはさまざまなピッチごとにデータを用意するのが望ましいが、このことによってデータベース中のデータの量はさらに大きくなってしまう。

【００４１】そこで、前記データベースのサイズを非常に小さくすることの出来る本発明の他の実施の形態について説明する。この実施の形態では、前記データベース１０を作成するときに、伸ばし音の非調和成分のデータを記憶する際、スペクトル包絡白色化手段を付加する。そして、合成時の前記非調和成分調整手段内に、非調和成分のスペクトル包絡生成手段を設けるようにしている。これにより、伸ばし音の非調和成分について、そのスペクトル包絡を個別に記憶する必要をなくし、データ量の削減を可能としている。

【００４２】図１１は、伸ばし音の場合における調和成分と非調和成分のスペクトル包絡の一例を示す図であ

る。この図に示すように、母音などの伸ばし音の場合の非調和成分のスペクトル包絡は、一般に、調和成分のスペクトル包絡に形状が似ている、すなわち、山や谷の位置がおおよそ一致している。したがって、調和成分のスペクトル包絡に何らかの操作（ゲイン調整、全体的な傾きの調整など）を行なえば、非調和成分のスペクトル包絡として妥当なものを得ることができる。また、伸ばし音では、対象区間内の各フレームでの各周波数成分の微妙なゆらぎが重要であり、このゆらぎの度合いは母音が変わってもさほど変らないと考えられる。そこで、非調和成分の振幅スペクトル包絡をあらかじめ何らかの形で平坦なものにして、もとの母音の音色の影響を取り去っておく（白色化する）。白色化により、見た目に平坦なスペクトルとされる。そして、合成時には調和成分のスペクトル包絡の形状をもとに非調和成分のスペクトル包絡を求め、前記白色化したスペクトル包絡にかけてやれば非調和成分の振幅スペクトル包絡を求めることができる。すなわち、スペクトル包絡のみ調和成分のスペクトル包絡をもとに生成し、位相についてはもともとの伸ばし音の非調和成分に含まれるものをそのまま利用する。このようにすることで、白色化された伸ばし音データをもとに、異なる母音の伸ばし音データの非調和成分を生成することが可能となる。

【００４３】図１２は、本発明のこの実施の形態における前記音韻データベース１０の作成処理を説明するための図であり、前記図１と同一の構成要素には同一の番号を付し、説明を省略することとする。この図１２に示すように、この実施の形態においては、伸ばし音について、前記区間切り出し手段１４から出力される非調和成分の振幅スペクトルを白色化するスペクトル白色化手段８０を有している。これにより、伸ばし音の非調和成分の振幅スペクトルとして白色化された振幅スペクトルのみが記憶されており、各素片データの非調和成分としてはその位相スペクトルのみが記憶されることとなる。

【００４４】図１３は、前記スペクトル白色化手段８０の一構成例を示す図である。前述のように、このスペクトル白色化手段８０により伸ばし音の非調和成分の振幅スペクトルは白色化され、見た目に平坦なものとされるのであるが、このときに、区間内の全てのフレームにわたってスペクトルを完全に平坦（全ての周波数で同一の値を持つ）にするのではなく、各周波数の時間的な微妙なゆらぎを残したまま各フレームの形状を平坦に近くする、という動作が必要になる。そこで、図１３に示すように、代表振幅スペクトル包絡作成部８１において、区間内の代表的な振幅スペクトル包絡を求め、スペクトル包絡の逆数生成部８２で、そのスペクトル包絡の各周波数成分の逆数を求め、これをフィルタ８３において、各フレームのスペクトル包絡のそれぞれの周波数成分に掛け算するという操作を行なっている。ここで、前記区間内の代表的な振幅スペクトル包絡を求めるには、例えば、各周波数ごとに平均値をとって、その平均値を使って代表的スペクトル包絡としてもよい。また、区間内の各周波数成分の最大値を使って代表的スペクトル包絡としても良い。これにより、前記フィルタ８３から白色化された振幅スペクトルが得られる。また、位相スペクトルはそのままその素片の非調和成分領域に記憶される。

【００４５】このようにして、伸ばし音の非調和成分は白色化されるが、合成時には調和成分のスペクトル包絡を用いて非調和成分を求めるので、白色化された非調和成分は、母音であればすべての母音に共通に使用することができる。すなわち、母音であれば、１つの伸ばし音の白色化された非調和成分があれば、充分である。もちろん、複数の白色化非調和成分を持っても差し支えない。

【００４６】図１４は、このように伸ばし音の非調和成分について白色化した振幅スペクトルを記憶するようにした場合の合成処理について説明するための図である。この図において、前記図２と同一の構成要素には同一の番号を付し、説明を省略する。この図に示すように、この実施の形態においては、前記音韻データベース１０から読み出された当該素片の非調和成分（白色スペクトル）が入力されるスペクトル包絡生成手段９０が、前記非調和成分調整手段２４の前段に付加されている。前述のように、前記音韻データベース１０から伸ばし音の白色化された非調和成分が読み出されたときは、スペクトル包絡生成手段９０において、調和成分のスペクトル包絡をもとに、非調和成分の振幅スペクトル包絡を計算する。例えば、最大周波数の成分は変わらないものとして、スペクトルの包絡の傾きだけを変えるように非調和成分のスペクトル包絡を定める方法が考えられる。そして、この振幅スペクトル包絡を同時に読み出された非調和成分の位相スペクトル包絡とともに、前記非調和成分調整手段２４に入力する。以下の処理は、前記図２に示した場合と同様である。

【００４７】このように、伸ばし音の非調和成分の振幅スペクトルを白色化して記憶する場合には、一部の伸ばし音についてのみ白色化された非調和成分の振幅スペクトルを記憶し、他の伸ばし音については非調和成分の振幅スペクトルを記憶しないようにすることができる。この場合には、合成時に、伸ばし音の素片データに非調和成分の振幅スペクトルがないので、合成する音素に最も近い音素をデータベース中から選択し、その伸ばし音の非調和成分を用いて、上述のようにしてその非調和成分の振幅スペクトルを作成すればよい。また、伸ばし音が可能な音素を１つ以上のグループに分け、合成する音韻が属するグループの伸ばし音データの内の１つを使用して、上述のように、非調和成分の振幅スペクトルを生成するようにしてもよい。

【００４８】なお、上述のようにして白色化された振幅スペクトルと調和成分の振幅スペクトルから求めた非調

19

和成分の振幅スペクトルを用いる場合に、その非調和成分の位相スペクトルの周波数軸の全てまたは一部を元データのピッチに対応する倍音付近のデータが再生する所望のピッチに対応する倍音付近に位置するように移動させる、すなわち、倍音付近の位相データは合成時にも倍音付近の位相データとして用いるようにすることにより、より自然な合成音とすることが可能となる。このようにこの実施の形態によれば、データベース中に全ての母音についての伸ばし音の非調和成分を記憶しておく必要がなくなり、データ量を削減することが可能となる。

【００４９】さらに、このスペクトルの包絡の傾きだけを変えることによって非調和成分のスペクトル包絡を定める場合に、その傾きの変化を「ハスキー度」と関連付けることにより、合成音声のハスキー度をコントロールすることができるようになる。すなわち、合成音声において非調和成分が多ければハスキーな声になり、少なければ潤いのある声になるので、傾きが急（０Ｈｚにおけるゲインが大きい）ならばハスキーな声になり、傾きがなだらか（０Ｈｚにおけるゲインが小さい）ならば潤いのある声になる。そこで、図１５に示すように、非調和成分のスペクトル包絡の傾きをハスキー度を表すパラメータで制御することにより、合成音声のハスキー度を制御することができる。

【００５０】図１６は、ハスキー度の制御を行なうことができるようにした場合の前記スペクトル包絡生成手段９０の構成例を示す図であり、スペクトル包絡生成部９１において、調和成分のスペクトル包絡に対して、コントロールパラメータとして供給されるハスキー度情報に従った傾きを乗算し、このようにして得られた特性を前記非調和成分の白色化された振幅スペクトルにフィルタ９２で付加する。そして、前記非調和成分の位相スペクトル包絡と前記フィルタ９２の出力を非調和成分のデータとして、前記非調和成分調整部２４に出力する。

【００５１】なお、調和成分のスペクトル包絡を何らかの形でモデル化し、その中のパラメータとハスキー度を関連付けても良い。例えば、調和成分のスペクトル包絡を定式化するときのパラメータのうちのいずれか（傾きに関連するパラメータ）を変化させることでハスキー度と関連付けて非調和成分のスペクトル包絡を求めても良い。また、ハスキー度は時間的に固定としても良いし、可変としても良い。可変にした場合、音韻を伸ばしている途中でだんだん声がハスキーになってくるというような面白い効果を得ることもできる。

【００５２】また、単にハスキー度の制御を行なうことができるようにするためには、上述のように、音韻データベース１０に非調和成分の白色化した振幅スペクトルを記憶しておく必要はない。上述した最初の実施の形態のように、伸ばし音の非調和成分についても他の素片と同様に振幅スペクトルをそのまま記憶しておき、合成時に、その非調和成分の振幅スペクトルに、その伸ばし音

20

区間内を代表する振幅スペクトルを求めてその逆数を乗算することにより平坦なスペクトルを作成し、調和成分の振幅スペクトルに基づいてハスキー度を制御するパラメータに応じた非調和成分の振幅スペクトルを計算し、前記平坦なスペクトルに乗算することにより得たスペクトルを非調和成分の振幅スペクトルとすればよい。

【００５３】

【発明の効果】以上説明したように、本発明の歌唱合成装置によれば、次のような効果を得ることができる。

・ＳＭＳ技術の利用により、了解度は良好で、伸ばしている部分も自然な合成歌唱音が得られる。

・ＳＭＳ技術の利用により、ビブラートやピッチの微妙な変化を行なった場合でも不自然な合成音にならない。

・有声音部分（調和成分）のスペクトル包絡の形状が最適なものを含む素片を選択あるいは補間により求めるため、ピッチによるスペクトル包絡の形状の変化にも対処することができる。その結果、幅広いピッチにおいて良い音色が得られる。

・有声音の場合の非調和成分について、所望のピッチに合うようにスペクトル形状の微細な形状を変化させるため、非調和成分と調和成分を混合しても雑音に聴こえたり浮いた音に聴こえたりすることがない。

・音素の伸ばし部分の長さや音素連鎖の長さを自由に調整できるので、所望のテンポどおりに合成歌唱音を得ることができる。

・音素・音韻の接続部分について、スムージング、あるいはその音素・音韻のレベル調整を行うため、接続時に雑音が発生しない。

・合成された歌声は、所望のピッチに合う音色になり、求めるタイミングで歌われ、接続単位間の雑音も無く、高い品質の歌声となる。

【００５４】また、伸ばし音の非調和成分を白色化して記憶する本願の歌唱合成装置によれば、データベースのサイズを非常に小さくすることができるとともに、データベース作成の効率を向上させることが可能となる。また、簡単に合成音声のハスキーさの度合いを調整することのできる歌唱合成装置を提供することが可能となる。

【図面の簡単な説明】

【図１】　本発明の歌唱合成装置に使用する音韻データベースの作成処理について説明するための図である。

【図２】　本発明の歌唱合成装置における歌唱音合成処理について説明するための図である。

【図３】　本発明の歌唱合成装置における非調和成分調整処理について説明するための図である。

【図４】　本発明の歌唱合成装置におけるループ処理について説明するための図である。

【図５】　スペクトル包絡のモデル化について説明するための図である。

【図６】　素片データの接続部におけるミスマッチにつ

いて説明するための図である。

【図７】　本発明の歌唱合成装置におけるスムージング処理について説明するための図である。

【図８】　本発明の歌唱合成装置におけるレベル調整処理について説明するための図である。

【図９】　本発明の歌唱合成装置の一実施の形態の構成を詳細に示す機能ブロック図である。

【図１０】　本発明の歌唱合成装置を動作させるためのハードウエア装置の一例を示す図である。

【図１１】　伸ばし音における調和成分と非調和成分のスペクトル包絡の一例を示す図である。

【図１２】　本発明の歌唱合成装置の他の実施の形態における音韻データベースの作成処理について説明するための図である。

【図１３】　スペクトル白色化手段の一構成例を示す図である。

【図１４】　本発明の歌唱合成装置の他の実施の形態における歌唱音合成処理について説明するための図である。

【図１５】　ハスキー度の制御について説明するための図である。

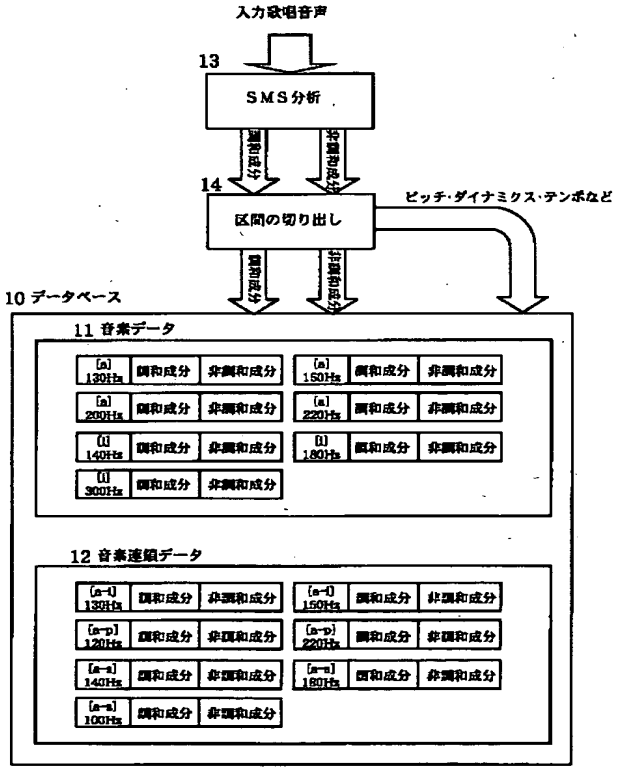
【図１６】　ハスキー度の制御を行なうことができるようにした場合のスペクトル包絡生成手段の構成例を示す図である。

【図１７】　従来のＳＭＳ方式を適用した歌唱合成装置について説明するための図である。

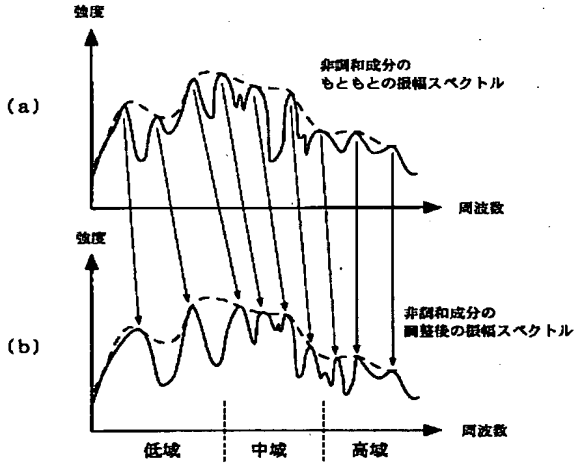
【符号の説明】

１０　音韻データベース、１３　ＳＭＳ分析手段、１４　区間切り出し手段、２１　音素→素片変換手段、２２　調和成分調整手段、２３　非調和成分調整手段、２４　継続時間調整手段、２５　素片レベル調整手段、２６　素片接続手段、２７　調和成分生成手段、２８　合成手段、８０　スペクトル白色化手段、９０　スペクトル包絡生成手段

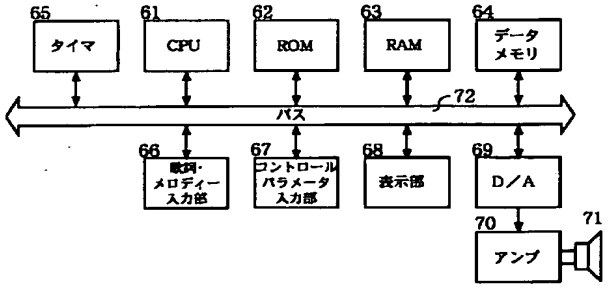
【図１】



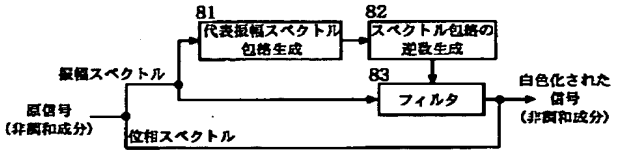
【図３】



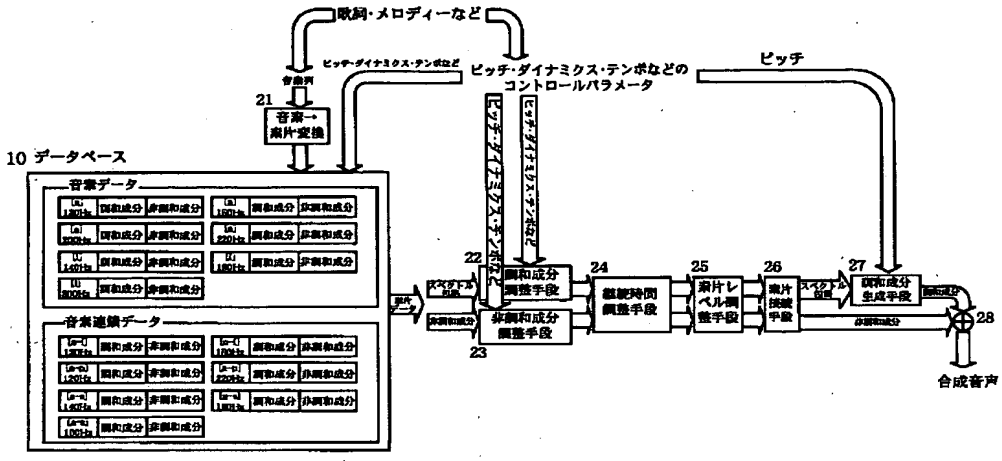
【図１０】



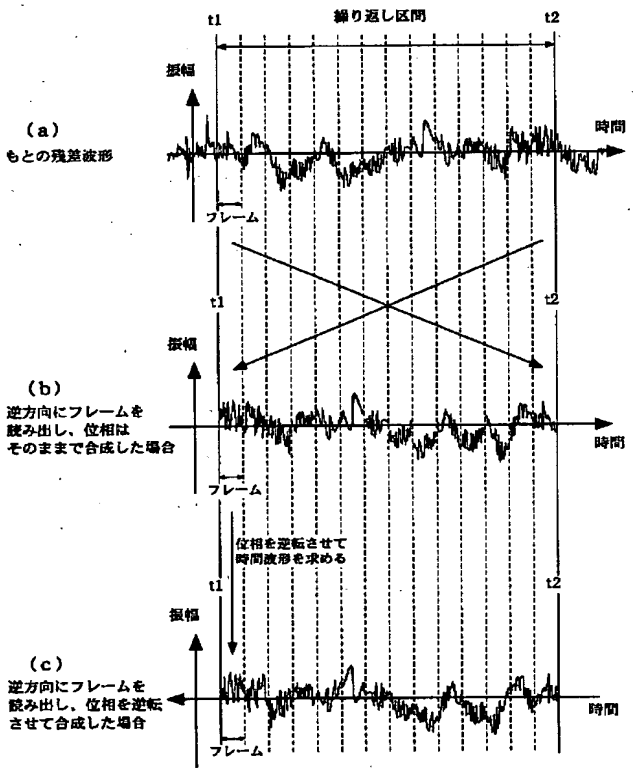
【図１３】



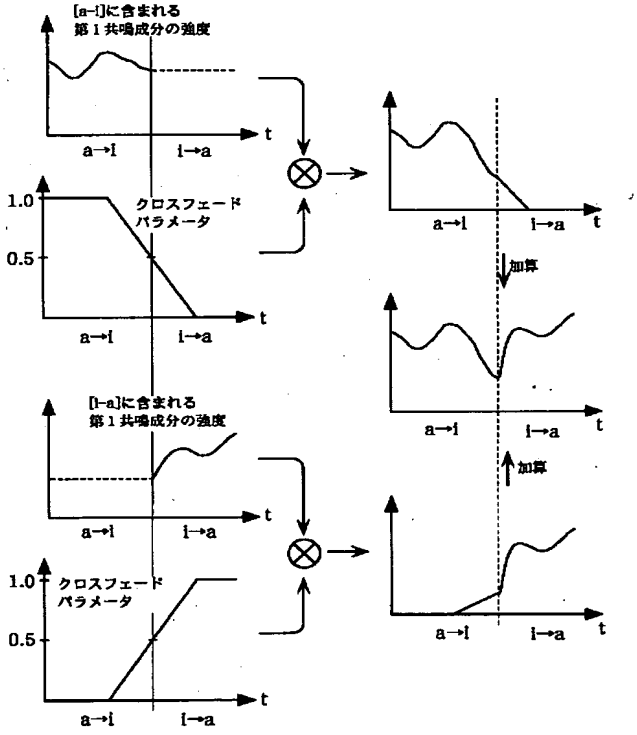
【図２】


歌詞・メロディーなど
ピッチ・ダイナミクス・テンポなど
音素列
21
音素→素片変換
ピッチ・ダイナミクス・テンポなどのコントロールパラメータ
ピッチ
10 データベース
音素データ
音素連鎖データ
22
調和成分調整手段
23
非調和成分調整手段
24
継続時間調整手段
25
素片レベル調整手段
26
素片接続手段
27
調和成分生成手段
28
合成音声

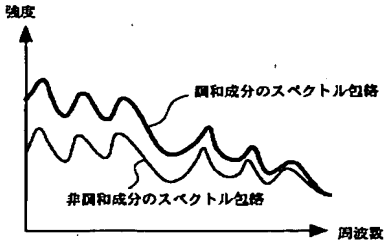
【図４】


t1
繰り返し区間
t2
振幅
時間
(a)
もとの残差波形
フレーム
(b)
逆方向にフレームを読み出し、位相はそのままで合成した場合
位相を逆転させて時間波形を求める
(c)
逆方向にフレームを読み出し、位相を逆転させて合成した場合

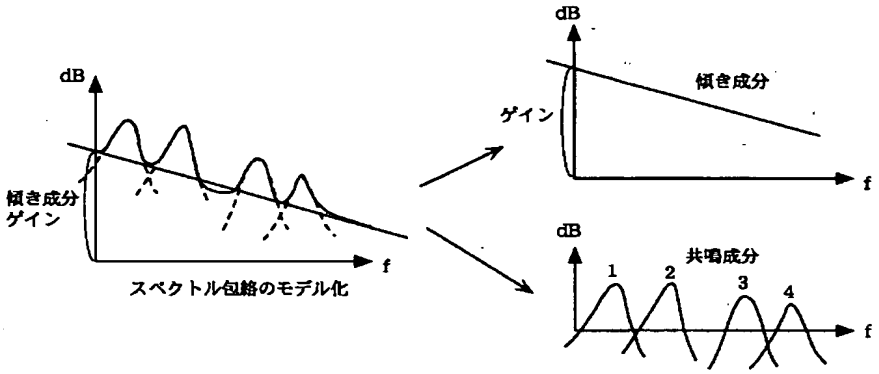
【図７】


[a-i]に含まれる第１共鳴成分の強度
a→i
i→a
t
1.0
0.5
クロスフェードパラメータ
加算
[i-a]に含まれる第１共鳴成分の強度

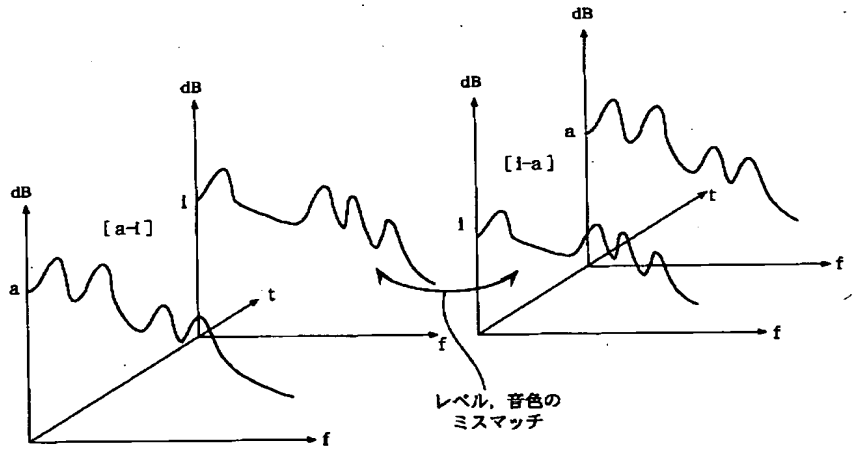
【図１１】


強度
調和成分のスペクトル包絡
非調和成分のスペクトル包絡
周波数

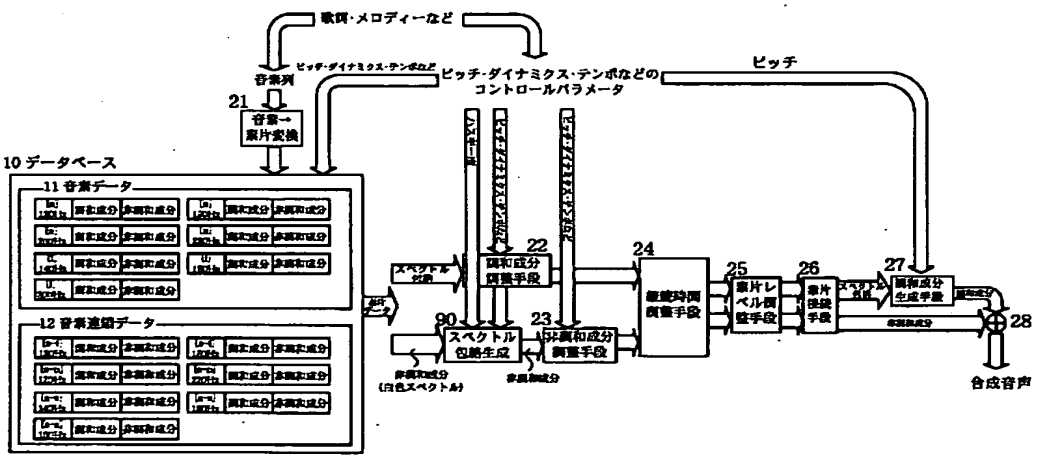
【図５】


dB
傾き成分
ゲイン
f
スペクトル包絡のモデル化
dB
傾き成分
ゲイン
f
dB
共鳴成分
1
2
3
4
f

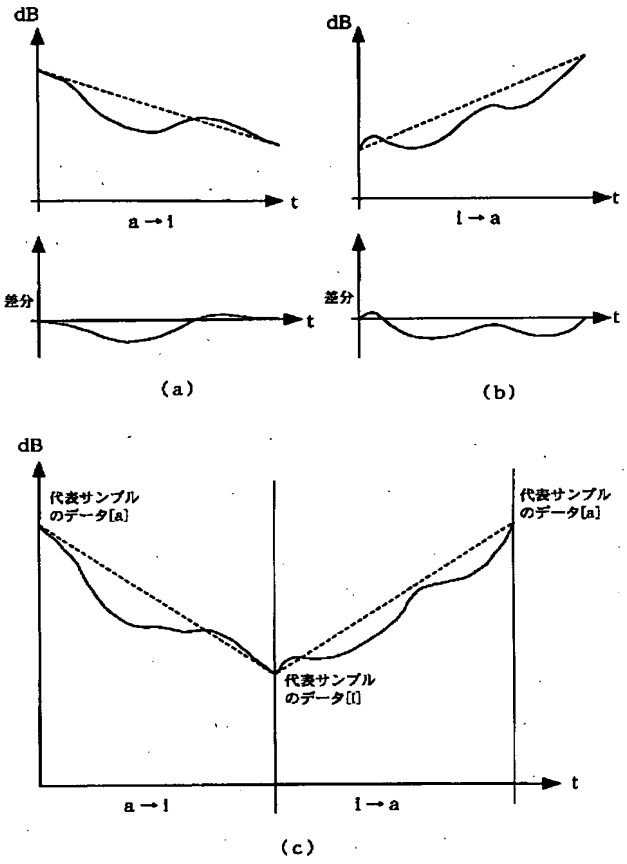
【図６】


dB
[a-l]
a
l
t
f
dB
[l-a]
a
l
t
f
dB
f
レベル，音色の
ミスマッチ

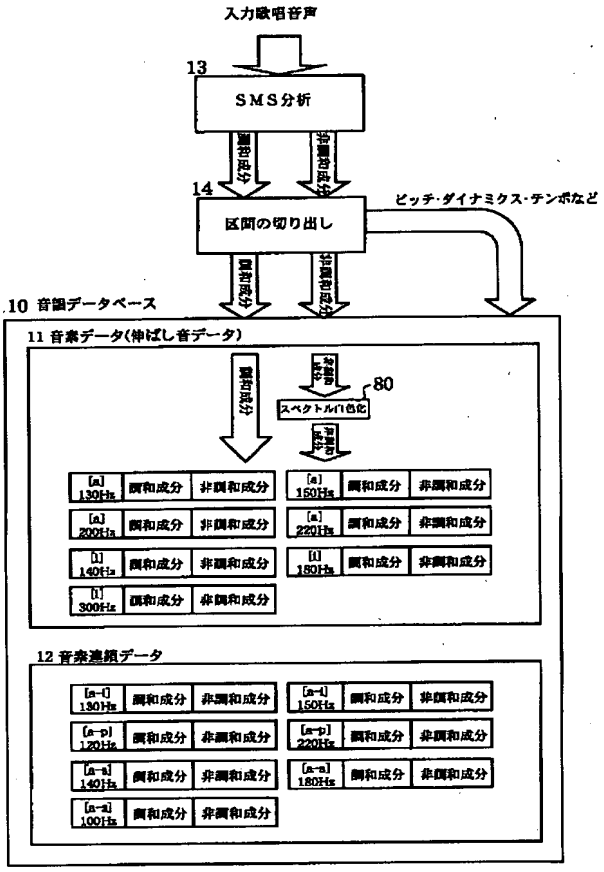
【図１４】


歌詞・メロディーなど
音素列
ピッチ・ダイナミクス・テンポなど
ピッチ・ダイナミクス・テンポなどの
コントロールパラメータ
ピッチ
21
音量→
声片変換
10 データベース
11 音声データ
12 音素連鎖データ
スペクトル
包絡生成
90
22
調和成分
調整手段
23
非調和成分
調整手段
24
継続時間
調整手段
25
声片レ
ベル調
整手段
26
声片
接続
手段
27
調和成分
生成手段
28
合成音声

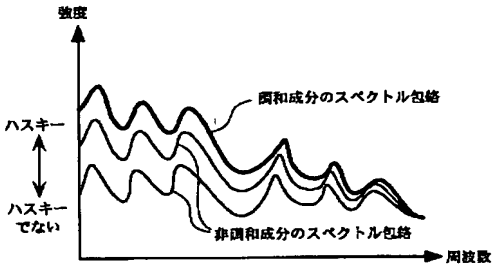
【図８】



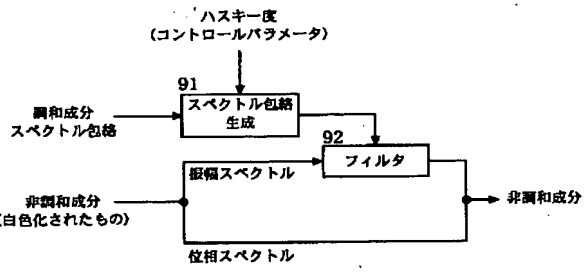
【図１２】



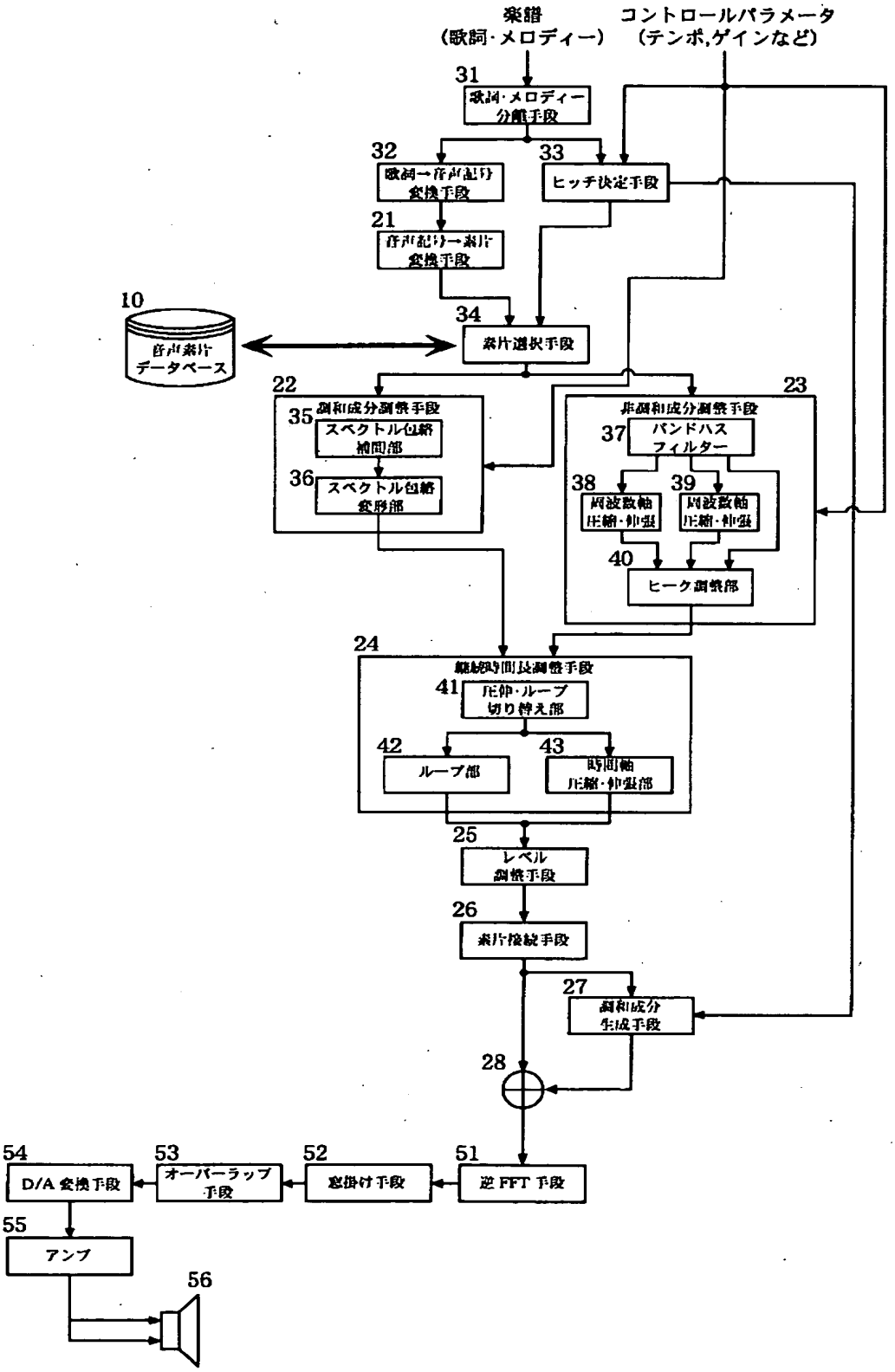
【図１５】



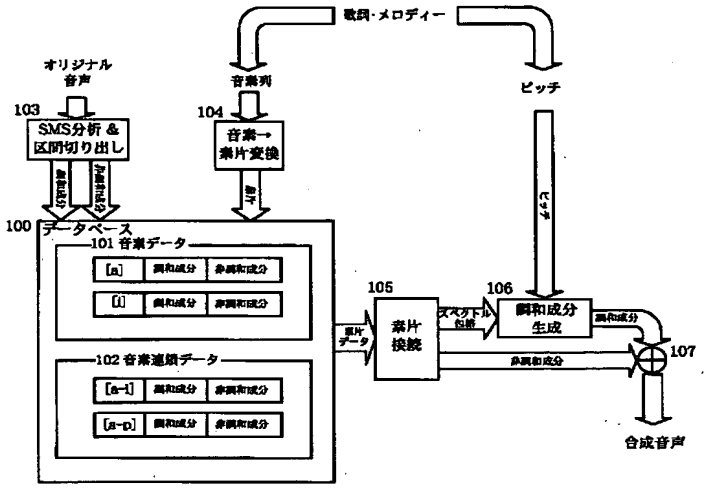
【図１６】



【図９】


楽譜
(歌詞・メロディー)
コントロールパラメータ
(テンポ,ゲインなど)
31
歌詞・メロディー
分離手段
32
歌詞→音声記号
変換手段
33
ピッチ決定手段
21
音声記号→素片
変換手段
10
音声素片
データベース
34
素片選択手段
22
調和成分調整手段
35
スペクトル包絡
補間部
36
スペクトル包絡
変形部
23
非調和成分調整手段
37
バンドパス
フィルター
38
周波数軸
圧縮・伸張
39
周波数軸
圧縮・伸張
40
ピーク調整部
24
継続時間長調整手段
41
圧伸・ループ
切り替え部
42
ループ部
43
時間軸
圧縮・伸張部
25
レベル
調整手段
26
素片接続手段
27
調和成分
生成手段
28
54
D/A 変換手段
53
オーバーラップ
手段
52
窓掛け手段
51
逆 FFT 手段
55
アンプ
56

【図１７】



フロントページの続き

(72)発明者　ジョルディ　ボナダ
スペイン　バルセロナ　08002　メルセ
12

Ｆターム(参考)　5D045 AA08 AA09